THE MAIN FEATURE FOR SUCCESSFUL IMPLEMENTATION A DATA WAREHOUSE

VISHAL KUMAR GOAR

Research Scholar, Suresh Gyan Vihar University,
Jaipur, Rajasthan, India
goar.vishal@gmail.com

PROF. (DR.) S.S.SARANGDEVOT

Director,J.R.N.Rajasthan Vidyapeeth,
Udaipur,Rajasthan, India
dr.sarangdevot@ gmail.com

Abstract

Data warehousing is the leading and most reliable technology used today by companies for planning, forecasting, and management A lot has been done in this field regarding design and development of data warehouses and a lot still needs to be done but one area which needs special attention from research community is data warehouse maintenance.

A major reason for data warehouse project failures is poor management. Without proper management desired results are nearly impossible to attain from a data warehouse. Unlike operational systems data warehouses need a lot more management and a support team of qualified professionals is needed to take care of the issues that arise after its deployment including data extraction, data loading, network management, training and communication, materialized view and some other related tasks. The subject of my research is to explore the impact of the selected factors, under organizational, project-related and environmental dimensions, on data warehouse applications.

***Keywords*:** Data warehousing; network management; training and communication; materialized view;

1. Introduction

Authors are encouraged to have their contribution checked for grammar. Abbreviations are allowed but should be spelt out in full when first used. Integers ten and below are to be spelt out. Italicize foreign language phrases (e.g. Latin, French). Over the years, firms have accumulated a large volume of data that potentially contain valuable information about their business. These data, which are stored in operational databases, are not easily accessible to decision makers due to most firms existing information technology (IT) infrastructure. Further, these operational databases are likely to be geographically or logically dispersed, thereby making it difficult and inconvenient to access.

Traditionally, the information systems Departments are the sole interface to firms' data stored in computer systems. Executives from various departments rely on the IS department to satisfy their needs for information that is necessary for decision making. The turnaround time is usually long and information is often not delivered on time to users, which in turn reduces the value of the information. Further, reports produced by the ARE departments usually pro-vide a one-dimensional (rather than multidimensional) view of the data.. Executives will require different reports showing information in various dimensions (e.g., by product lines or by region.). Further, the ability to change dimensions and drill down enables more effective data analysis and decision making. However, since senior executives tend to ask for whatever they need, it is usually time consuming for the IS departments to deliver different reports customized to different executives. Consequently, only a subset of

reports may be delivered on time. Even if the reports are delivered on time, it may not be exactly what the executive wants due to miscommunication or lack of clarity in specifying the information he/she needs. Further, on receiving the reports, the executive may discover that he/she needs further detailed information, which in turn may require more time for the IS department to provide. Even if further information is not needed, the executive may still need to seek clarification from the IS department regarding various aspects of the information provided. This is likely to result in further delay, thereby hindering timely and effective decision making.

With the appropriate user friendly query tools, staff members can experiment with different views of the data, thereby enabling him/her to understand the situation and make better decisions. Further, if necessary, staff members can drill down to obtain more detailed data about certain areas. This ability to change views and drill down greatly enhances the value of a data warehouse for decision making since detailed data can be readily accessed, whenever required. Past research has often focused on the technology required to establish and support data warehouses [14]. Previous work has also provided both technical [16] and non-technical prescriptions [18] for building and implementing data warehouses. The factors affecting data warehouse implementation [14], warehouse issues [13, 15], as well as the pitfalls of data warehouse development [19] have also been investigated. In addition, comparative cases of data warehousing [13] and businesses, which successfully deploy data warehouse, are often described.

The text is to be typeset in 10 pt roman, single spaced with baselineskip of 13 pt. Text area is 5 inches in width and the height is 8 inches (including running head). Final pagination and insertion of running titles will be done by the publisher. Upon acceptance, authors are required to submit their data source file including postscript files for figures.

2. Performance Effect Factor

While the implementation of a specific phase of the data warehouse may be completed, but the data warehouse program needs to be continued [3]. Progress monitoring needed to be continued against the agreed on success criteria. The data warehouse team must ensure that the existing implementations remain on track and continue to address the needs of business. Performance issues in data warehousing are centralized around access performance for running queries and incremental loading of snapshot changes from the source systems [4,5,6] The following six concepts can be considered for a better performance:

2.1. *Valuable Communication*

A Data ware house system is actually about strongly combine different business functions, so the close co-operation and communication across disparate business functions would be a natural prerequisite in a Data ware house project. Some authors have described the co-ordination and communication between departments as the oil that keeps everything working properly in these contexts. [7, 8]

An effective communications program keeps people informed and generates interest in using the data warehouse. Organizational rules to rationalize inconsistencies had to be established. Clear and consistent communication of company-wide warehouse goals and policies fosters employee participation on three critical fronts: First, it reinforces the front-line employees' contribution of information to the warehouse. Second, it encourages information sharing to support ongoing business activities. Third, it inspires mid-level managers to use the data warehouse to inform key stakeholders regarding decisions, and new projects.

The cooperation between the departments in an organization has a large effect on the smooth flow of the required information and expertise among the departments, which strongly influences the successful adoption of data warehouse technology. Interdepartmental cooperation and communication is a must for any project.

2.2. *Education, Training and Documentation*

Training and education of the employees are required in a successful data warehouse project. A data warehouse is not a simple project or an easy-to-learn system. It demands time to educate and transfer the knowledge to users by setting up training courses and distributing related-material. In most computing projects, management identifies the need for training, but does not always fund training. With every new database there is a need for

another training course, complete with reference materials. Every enhancement or change to the warehouse must be documented and communicated to warehouse users.

Training expands the communication process by maintaining a level competence in both the business and IT community as to the tools and mechanisms of the data warehouses. The quality of employees and their development through training and education are major factors in determining long-term profitability of a small business. If you hire and keep good employees, it is good policy to invest in the development of their skills, so they can increase their productivity. Training often is considered for new employees only.

This is a mistake because ongoing training for current employees helps them adjust to rapidly changing job requirements. Training and updating the employees' knowledge of data ware house is a major challenge. Data ware house implementation requires a huge mass of knowledge to enable people to use, cope and solve problems within the framework of the system. Training employees to use ERP is not as simple as training them in any other packaged-software such as a Microsoft package.

An ERP system is extremely complex and demands intensive training; it is difficult for the trainers to pass the knowledge to the users within a short period of time. Top management should understand this aspect and should be willing to spend adequate money on educating and training the end users. [9]

2.3. *Call Center*

Another role of the data warehouse support and protection group is the problem resolution when some problem is encountered in the data warehouse. In the case of data warehouses, the expensive and the risky nature of data warehouses have forced the potential adopters to pay extra attention in selecting appropriate vendors to increase the possibility of having successful data warehouse initiatives.

The Call Center is an important division for any organization as it serves as the primary interaction point between customers and the company. In many situations, it is the only interaction point and therefore, responsible for the customer's experience and satisfaction. Due to this heightened level of importance, it is critical that the contact handling process is conducted both efficiently and effectively. This process specifies how to collect, document, answer and/or escalate calls, requests, and queries related to issues with the data warehousing environment. Problem documentation can be completed either by the Call Center representative and/or in conjunction with a form completed by the end user or IT support person requesting a service or action. All inquiries, no matter how trivial should be logged, especially during the start of a new data warehouse or mart. These bits of information can form clues to taking proactive action to bigger problems before they emerge.

2.4. *Network Management*

If there is a heterogeneous group of platforms for the data warehouse implementation, network management is going to be one of the most demanding tasks [6]. Modern communication networks create large amounts of operational data, Including traffic and utilization statistics and alarm/fault data at various levels of detail. These massive collections of network-management data can grow the order of several Tera bytes per year, and typically hide "knowledge" that is crucial to some of the key tasks involved in effectively managing a communication network (e.g., capacity planning and traffic engineering).

Besides providing easy access to people and data around the globe, modern communication networks also generate massive amounts of operational data throughout their life span. As an example, Internet Service Providers (ISPs) continuously collect traffic and utilization information over their network to enable key network- management applications.

Not only are users coming constantly on-line, but users and equipment are invariably moving to new locations. The networking hardware is proliferating with LANs, WANs, hubs, routers, switches and multiplexers. Leaving behind all this is the next stage – users wanting to access internet based data sources

along with the corporate data, requiring even greater bandwidth and network management resources. Managing this environment is one big challenge, capacity planning for the future is another. If the data warehouse team is not quite good in networking technology than there should be at least one person in the organization who understands technology.

2.5. *Extract Transform and Load (ETL)*

The ETL process is much more than code written to move data. The ETL architect also serves as the central point for understanding the various technical standards that need to be developed if they don't already exist. These might include limits on file size when transmitting data over the company intranet, requirements for passing data through firewalls that exist between internal and external environments, data design standards, standards for usage of logical and physical design tools and configuration management of source code, executables and documentation. The ETL architect must also ensure that the ETL design process is repeatable, documented and put under proper change control.

Extract, transform and load (ETL) is the core process of data integration and is typically associated with data warehousing. ETL tools extract data from a chosen source, transform it into new formats according to business rules, and then load it into target data structure. [VG10] ETL which stands for extract, transform, and load is a three-stage process in database usage and data warehousing. It enables integration and analysis of the data stored in different databases and heterogeneous formats. After it is collected from multiple sources (extraction), the data is reformatted and cleansed for operational needs (transformation). Finally, it is loaded into a target database, data warehouse or a data mart to be analyzed.

A key consideration for the ETL architect is to recognize the significant differences that the design and implementation methods for a business intelligence system have from an online transaction processing (OLTP) system approach. One last role for the ETL architect must be to ensure that the various software tools needed to perform the different types of data processing are properly selected ETL is one of the most important sets of processes for the sustenance and maintenance of Business Intelligence architecture and strategy [12].

If source data taken from various sources is not cleanse, extracted properly, transformed and integrated in the proper way, the extracted data will often be stored in a central staging area where it will cleanse and otherwise transformed before loading into the warehouse. An alternative approach to information integration is that of mediation: data is extracted from original data sources on demand when a query is posed, with transformation to produce a query result

2.6. *Materialized View*

A view can be materialized by storing the tuples of the view in the database. Index structures can be built on the materialized view. Consequently, database accesses to the materialized view can be much faster than recompiling the view. A materialized view is thus like a cache {a copy of the data that can be accessed quickly.} a materialized view provides fast access to data; the speed difference may be critical in applications where the query rate is high and the views are complex so that it is not possible to recomputed the view for every query.

Materialized views provide a framework within which to collect information into the warehouse from several databases without copying each database in the warehouse. Queries on the warehouse can then be answered using the materialized views without accessing the remote databases. Provisioning, or changes, still occurs on the remote databases, and are transmitted to the warehouse as a set of modifications. Incremental view maintenance techniques can be used to maintain the materialized views in response to these modifications.

Materialized views, and many recognize dependencies between base tables and the Materialized on of very efficient refresh expressions. This makes Database Server more manageable and user friendly

3. Conclusion

A data warehouse solution is not only a software package. It is a complex process to establish sophisticated and integrated information systems. The adoption of this technology requires massive capital expenditure, utilizes a certain deal of implementation time and has a very high likelihood of failure. Therefore, many adoption-related factors must be carefully assessed before the real adoption is actualized.

I do study on BSNL its shows that the first and the most important part of a data warehouse maintenance program is the training of its users. The study shows that most business users are reluctant to adopt technology to carry out their work, therefore pursuing a business user to use data warehouse is inevitable. To pursue business users in using data warehouse the communication and training program is a must. The training program gives the users of data warehouse an insight into the qualities and capabilities of a data warehouse and teaches them

The communication process keeps the business users and IT users in contact with each other to have exchange of views, suggestions and any guidance towards enhanced performance of a data warehouse. The services of help desk and problem management play an important role in taking valuable output from the data warehouse. Support is always required in any information system, same is the case with a data warehouse but here the support is needed 24 hours a day. Some of the processes like ETL are carried out during the night, which require presence of support staff to rectify any problem.

The support team also points out if there is any loop hole or problem area within the data warehouse that should be addressed. Apart from help desk each data warehouse support team develops its own problem management process. The process defines necessary routines and instructions to counter any problem found in the warehouse. If the problems found in the data warehouse are not addressed at the right time, this leads to performance shortfalls, and usability and availability issues in near future. Thus help desk and problem management play a key role in improving data warehouse performance and getting the desired out put from it.

Network management also plays its part in improving data warehouse performance. From the case study we concluded that by having a fast and reliable network user queries get a much shorter response time especially in a distributed data warehouse.

The ETL architect has a clear understanding of the company's business and knows what type of data is exactly required. The data warehouse will not address reporting requirements, until and unless it has the data that is useful. If the data is of no use for the business, there is no need storing it in the warehouse.

View materialization is a strategy used to provide fast answers to user queries. But it is important to have updated views whenever the base tables upon which views are built are updated. It is the responsibility of data warehouse support team to devise a flexible and optimal strategy for maintenance of materialized views. Although it is more of a design decision to provide capabilities for view maintenance but it is the user's decision to decide which views to materialize and when to refresh them.

References

[1] An investigation of the factors affecting data warehousing success by Roger L. Hayen et. al. in journal of Issues in Information Systems, Volume VIII, No. 2, 2007.

[2] Developing a Data Warehouse Process that responds to the needs of the Enterprise,Peter R. Welbrock Smith-Hanley Consulting Group Philadelphia, PA

[3] The data warehouse toolkit. 2nd edition. Ralph Kimball, Margy Ross. 2002 Wiley computer publishing..

[4] Data Warehouse Management Handbook by Richard Kachur. 2000 Prentice Hall .

[5] Building, using, and managing the data warehouse. Ramon Barquin, George Zagelow,Katherine hammer, Mark sweiger, George Burch, Dennis Berg, Christopher Heagele,Katherine Glassey-Edholm, David Menninger, Paul Barth, J.D. Welch, Narsim ganti, Herb Edelstein, Bernard Boar, Robert Small. Data warehousing institute series from Prentice Hall.Y.

[6] Data warehouse, Practical advice from the experts. 1997. Prentice hall by Joyce Bischoff & Ted Alexander

[7] Akkermans and Helden, Vicious and virtuous cycles in ERP implementation: A case study of interrelations between critical success factors, European journal of information systems, 2002, Vol.11 Iss. 1, p35..

[8] Nah et al., Critical factors for successful implementation of enterprise systems, Business process management, Bradford: 2001, Vol.7 Iss.3, p285.:

[9] Bingi et al., Critical issues affecting an ERP implementation, Information systems management, 1999, Vol.16 Issue3.

[10] Lessons from a successful data warehouse implementation. Dr. John D Porter and john. J Rome. Arizona State University.

[11] Building a data warehouse for decision support. 1996 Prentice Hall. By Vidette Poe with contributions from Laura L. Reeves.

[12] Fundamentals of database systems. 4th Edition. Persons international and Addison Wesley. Ramez Elmasri and Shamkant B. Navathe

[13] N. Alur, Missing links in data warehousing, Database Programming and Design _1995. 21–23, September.

[14] E. Appleton, The right server for your data warehouse,Datamation 41 _5. _1995. 56–58, March.

[15] S. Atre, Rules for data cleansing, Computerworld _1998.69–72, March 8.

[16] D.P. Ballou, G.K. Tayi, Enhancing data quality in data warehouse environments, Communications of the ACM 42 (1) (1999) 73–78.

[17] J. Bischoff, Achieving warehouse success, Database Programming and Design (1994) 27–33, July.

[18] S. Deck, Data warehouses: plan well, start small, Computer world _1998. 9, August 3.

[19] J. Foley, Data warehouse pitfalls, Informationweek _1997. 93–96, May 19.